BIOSENSOR MATERIALS AND METHODS

Technical Field

This invention relates to biosensor materials and methods, and in particular to methods for generating microorganisms having utility in biosensing, tools which can be generally used in such methods, the microorganisms themselves, and biosensing methods employing such microorganisms.

Background Art

It is frequently desirable to be able to detect small concentrations of analytes in samples, e.g. environmental samples. For instance, to allow more effective management of scarce environmental resources, more efficient and faster methods of assessing environmental pollution are required. At present, molecular-specific monitoring of effluent streams and other environmental matrices requires extensive chemical manipulation of the sample followed by Gas Chromatography (GC) and Mass Spectrometry (MS) analyses. Although these techniques are highly sensitive, sample preparation is necessarily slow and expensive. Consequently, continuous on-site analysis of a variety of environmental matrices cannot be achieved using these methods at reasonable cost.

An alternative method for the determination of phenols and chlorophenols has been proposed using a biosensor based around *Rhodococcus* sp. [see Riedel et al (1993) Appl Microbiol Biotechnol 38: 556-559]. In this method microorganisms are immobilised in an oxygen electrode, and oxygen uptake in response to added substrates was monitored. Although fairly simple and rapid, this method lacks robustness and is not sufficiently sensitive or specific for detecting particular environmental pollutants.

It can thus be seen that the provision of novel

materials and methods capable of being used in the field of biosensing would represent a step forward in the art.

Disclosure of Invention

In a first aspect of the invention there is disclosed a method of detecting the presence or absence of an analyte in a sample comprising the steps of:
(a) contacting the sample with a transformed microorganism which is a mycolic acid bacterium which expresses a binding agent capable of binding the analyte, wherein the binding of the agent to the analyte causes a detectable signal, and wherein said bacterium has been transformed such as to improve the detectability of the signal; and
(b) observing said bacterium for said detectable signal;

By "observing" is meant ascertaining by any means (directly or indirectly) the presence or absence of the selected signal which is indicative of the binding event.

By "improve" is meant, inter alia, altering the nature of the signal to one which can be observed more readily or increasing the intensity of the signal (thereby reducing the sensitivity of the means used to observe it).

Thus by using a transformed microorganism, the limitations inherent in wild-type microorganisms such as those used in the prior art may be overcome. In particular more sensitive and robust monitoring methods than those based on natural biochemical activities such as oxygen uptake can be employed. The mycolic acid bacterial gene expression-based sensors of the present invention can combine high sensitivity with the biofiltering and bioconcentrating aspects of the mycolic acid bacterial cell wall. Methods for generating such transformants are described in further detail below. Such transformed microorganisms are hereinafter referred to as 'biosensors'.

Preferably the analyte is an environmental

pollutant, for instance such as may result from industrial or medical applications. Of particular interest is the detection of mono- and poly-aromatic, cyclic, heterocyclic and linear hydrocarbons such as, but not limited to, components of fuels, solvents, propellants, energetics and pesticides (such as may appear on United States EPA Priority Pollutants List and European Community Grey and Black Lists) and naturally occurring degradation products of these compounds in industrial process media, vapours, effluents, raw water, rivers, ground waters, or soils. As will be clear to the skilled person from the disclosure hereinafter, the methodology of invention is inherently flexible and may, in principle, be employed to develop mycolic bacteria capable of biosensing almost any target analyte.

The mycolic acid bacteria form a supra generic group of Gram-positive, non-sporulating bacteria which is comprised of the genera Corynebacterium, Mycobacterium, Nocardia, Rhodococcus, Gordona, Dietzia and Tsukamurella. Members are metabolically diverse and capable of using as sole carbon source (a growth-inducing substrate) a wide range of natural and xenobiotic compounds, including many key environmentally-toxic and/or industrially-important molecules e.g. hydrophic organic compounds. The mycolic acid bacteria exhibit several structural and physiological features which appear to be specialisations for hydrocarbon degradation, these include a hydrophobic mycolic acid outer cell layer and associated production of extracellular mycolic acid-derived biosurfactants. Most preferably the bacterium is a member of the Rhodococcus or Nocardia complex (i.e. nocardioform actinomycete).

The detectable signal may be a change in enzyme function(s), metabolic function(s) or gene expression.

Preferably however the signal is ascertained in consequence to an increased expression of a signal protein from a signal gene, more preferably a

heterologous signal gene. Many suitable signal proteins (which have a readily detectable activity) are known in the art e.g. ßgalactosidase, which can generate a coloured substrate. The signal may utilise co-factors. Most preferably the activity of the signal protein, or the protein itself, can be estimated photometrically (especially by fluorimetry). This may be directly e.g. using instance green (and red) fluorescent protein, insect luciferase, and photobacterial luciferase. Alternatively it may be indirect e.g. whereby the signal gene causes a change which is detected by a colour indicator e.g. a pH change. Methods for introducing signal genes into appropriate hosts are described in further detail below.

Generally the bound agent/analyte complex will initiate expression of a signal gene which is operably linked to an inducible promoter. The identification of suitable promoters and/or coding sequences which are operably linked to them (including that of the binding protein) in mycolic acid bacteria, in order to modify said suitable promoters and/or coding sequences to introduce signal genes therein forms one part of the present invention.

As used herein, "promoter" refers to a non-coding region of DNA involved in binding of RNA polymerase and other factors that initiate or modulate transcription from a coding region of DNA whereby an RNA transcript is produced.

An "inducible" promoter requires specific signals in order for it to be turned on or off.

The terms "operatively linked" and "operably linked" refer to the linkage of a promoter to an RNA-encoding DNA sequence, and especially to the ability of the promoter to induce production of RNA transcripts corresponding to the DNA sequence when the promoter or regulatory sequence is recognized by a suitable polymerase. The term means that linked DNA sequences (e.g., promoter(s), structural

gene (e.g., reporter gene(s)), terminator sequence(s), are operational or functional, i.e. work for their intended purposes.

As is known to those skilled in the art, the transport and binding proteins (agents) required for the functionality of the inducible promoter, as well as the catabolic enzymes induced by it, will frequently form part the operon containing the promoter, and may thus be identified and isolated along side it using the methods disclosed above. These additional proteins are hereinafter referred to as "operon proteins".

Generally speaking, those skilled in the art are well able to construct vectors and design protocols for recombinant gene expression in common hosts such as E. coli. Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. For further details see, for example, *Molecular Cloning: a Laboratory Manual:* 2nd edition, Sambrook *et al*, 1989, Cold Spring Harbor Laboratory Press. Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in Molecular Biology*, Second Edition, Ausubel et al. eds., John Wiley & Sons, 1992. The disclosures of Sambrook et al. and Ausubel et al. are incorporated herein by reference.

However, the present inventors have recognised that certain methods previously employed in the art which were developed for enteric bacteria such as E. coli may not be the most appropriate for use in mycolic acid bacteria. The mycolic acid layer and associated biosurfactants (which are a defining feature of these bacteria) and thick cell wall confer great resistance to cell lysis

protocols known in the art. Similarly, mycolic strains used in the invention may not (indeed generally will not) be laboratory type strains, and may thus exhibit very high levels of nuclease activity.

In addition the detailed chemistry of the inducible pathway which forms the basis of the biosensors of the present invention will frequently not be known e.g. if there are no known enzyme pathways leading to the degradation of a particular analyte, or possibly the analyte is not mineralised completely and is only partially utilised in an uncharacterised but inducible pathway. Therefore cloning by acquisition of some defined enzyme activity, assayed through a particular reaction (as opposed to a general phenotypic activity which results in gain of utilisation of a particular analyte as a source of metabolically useful products) may not be a plausible option to isolate genes from a wild type mycolic acid bacterium.

Accordingly, advantageous methods have been developed by the inventors which in preferred forms allow the rapid isolation and characterisation of promoters and operably linked operon proteins which avoid or at least minimise host restriction and requires no prior knowledge of the inducible enzyme chemistry involved. The methods of identifying, modifying and employing novel inducible promoters and/or coding regions operably linked to them which are appropriate to mycolic acid bacteria are detailed below.

Thus in a second aspect of the invention there is disclosed a method for identifying DNA encoding an inducible promoter which is induced in response to a specific analyte and/or identifying DNA encoding associated operon proteins comprising the steps of:

(a) culturing a source of mycolic acid bacteria in a selective medium containing said specific analyte and being selective for oligotrophic bacteria,

(b) identifying bacteria capable of subsisting on said

medium,

(c) extracting DNA from said bacteria

(d) incorporating said DNA into vectors

(e) cloning said vectors into a suitable host cells

(f) screening the host cells for said inducible promoter and/or proteins in order to identify vectors encoding it.

By "screening" is meant subjected to analysis in order to determine the presence or absence of a particular defined property or constituent. Generally, in order to construct a biosensor strain against a particular analyte, isolation *de novo* from the soil or other environmental matrices of mycolic acid bacteria which exhibit inducible expression of catabolic genes in the presence of the analyte will be required. Methods of screening are discussed in more detail below.

As is known to those skilled in the art "oligotrophic bacteria" are bacteria which exhibit a preference for, and persistent slow growth on, very low levels of bioavailable carbon sources. These bacteria are adapted to and predominate in carbon-poor environments (predominantly aquatic habitats where carbon is limiting to μM levels). The term as used herein is intended also to embrace those bacteria which are capable of growing on defined minimal media without supplementary amino acids and vitamins (sometime termed prototrophic). Such bacteria are rarely capable of the very rapid growth as exemplified by the enteric bacterium E. coli, but are by contrast, extremely persistent and metabolically versatile. Work done by the present inventors has shown that, generally speaking, auxotrophic bacteria are not suitable as biosensor strains for environmental and industrial use.

Preferably the medium used in the second aspect is a defined minimal medium called hereinafter 'MMRN' which has been developed by the present inventors to screen for the oligotrophic mycolic acid bacteria (especially rhodococcal and nocardial strains) which form the basis

of the biosensor. This medium preparation is a derivative of von der Osten et al.(1989) but for mycolic acid bacteria sodium citrate and biotin are omitted. Most importantly, the level of carbon supplement is reduced to oligotrophic levels (<500 μM, more preferably <100 μM). Experiments show that MMRN facilitates simple, selective enrichment for oligotrophic, mycolic acid-containing bacteria as well as providing the basis for testing and characterisation of gene induction. The medium forms a third aspect of the present invention.

DNA may be extracted from the bacteria by any methods known in the art. However, the present inventors have demonstrated that DNA isolation from mycolic acid soil bacteria (particularly novel isolates which are generally highly resistant to lysis) using standard techniques is inefficient. Accordingly, several optimised methods of generating total DNA from mycolic bacteria have been developed, as described in more detail below (Examples 3 and 4). These involve bacterial culture in MMRN supplemented with L-glycine, oligotrophic levels of carbon source (80 μM) and removal of biosurfactants by washing in a non-ionic detergent (e.g. Tween 80) prior to a modified alkaline lysis technique. The concept of using a non-ionic detergent at between 0.05 - 0.5 % (preferably 0.1%) in order to facilitate DNA extraction is central to the novel, optimised methods.

"Vector", unless further specified, is defined to include, inter alia, any plasmid DNA, lysogenic phage DNA and/or transposon DNA, in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform a prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g. autonomous replicating plasmid with an origin of replication).

Preferably the host used is E. coli. More preferably it is an E. coli strain carrying one or more of the

mcrABC mrr hsdSRM recA and recO mutations, since this is believed to enhance clone recovery when using DNA derived from mycolic acid bacteria which (e.g. in Rhodococcus/Nocardia) is GC rich,. Gene libraries may be readily maintained in these strains.

Preferably the vector used with E. coli further incorporates the 'cos' element (which is well known to those skilled in the art). Because of their capacity and selection for large DNA inserts and efficient transfection rates, cosmid cloning vectors facilitate rapid gene library construction, which is especially useful in the present context because the activities of interest are often encoded by closely lined genes or operons which may be contained on relatively large fragments of the e.g. Rhodococcus/Nocardia genome.

Preferably the mycolic acid bacteria isolates are further screened, for instance after stage (b), to ensure an absence of catabolic repression. Catabolite repression is the selective control of gene expression in response to the energy state of the cell. This process is part of a range of gene expression strategies grouped under the "stringent/relaxed" responses. Together, these allow bacteria to optimise their metabolism for maximum energy efficiency. At the genetic level, catabolite repression is achieved by the selective expression of one of several sigma factors, each expressed under a different physiological state and/or growth phase (Fujita *et al*, 1994) each recognising a different promoter sequence (Bashyam *et al*, 1996). This facilitates the selective expression or repression of a wide range of genes and operons simultaneously via the regulation of a single gene product.

To create an efficient, functional biosensor, such media-associated repression/activation phenomena must be absent or be disabled in the host strain since, in principle, catabolite repression could seriously compromise the activity of a biosensor because the

presence of a more efficient carbon source (such as glucose, succinate or acetate etc.) would lead to repression of hydrocarbon catabolic pathways which forms the basis of the sensor. Mycolic acid bacteria Brevibacterium (Oguiza *et al.*, 1996), Corynebacterium, Nocardia (Takahashi, *et al*, 1991), Mycobacterium smegmatis, M. Tuberculosis, and M. bovis BCG (Bashyam *et al*, 1996) and M. leprae (Doukhan *et al*, 1995) encode multiple sigma factor genes consistent with global stringent/relaxed genetic control. Consistent with these data, catabolite repression has been experimentally observed in Rhodococcus (Baryshnikova, *et al*, 1997).

To identify strains lacking catabolic repression, the concentrations of an enzyme known to be, or suspected of being, associated with the catabolic pathway of interest (e.g. catechol 2,3-dioxygenase, which is associated with toluene catabolism) is assessed in (a) selective medium supplemented with the specific analyte, (b) selective medium supplemented with the specific analyte plus a high efficiency carbon source such as glucose (1 mM) and (c) selective medium supplemented with glucose (1 mM) alone. Enzyme activities should be very low or undetectable in the absence of analyte. In the presence of analyte, and glucose plus analyte, the activities should be, within experimental error, very similar. To ensure that not only are biosensor strains free from all complex media-associated repression/activation effects, microbiological screenings are preferably extended to include several complex media. e.g. Lauria Bertini broth or Nutrient Agar in addition to MMR + 1mM levels of individual carbon sources.

The present inventors have established that catabolic genes in mycolic acid bacteria exhibit poor DNA sequence conservation with analogous enzyme genes in Gram negative bacteria. As a result, "reverse genetic" approaches to isolation of novel catabolic pathways are likely to be of limited use when using such published

sequence data.

Thus in one embodiment of the second aspect, the host cells are screened for the inducible promoter and/or operon proteins by screening the cells using one or more probes based on the sequence of other promoters and/or operon proteins employed by mycolic acid bacteria in catabolic enzyme production. One example of a source of suitable sequences is the promoter operator region of the R. corallina orthohydroxyphenylpropionic - ohp - acid catabolic operon (which we had previously designated the monoaromatic catabolic - mac - operon) the sequence of which has been made available by the present inventors for the first time. This is described in more detail below, and in Example 9. Thus an inducible promoter and/or operon proteins may be identified by providing a nucleic acid molecule having a nucleotide sequence identical to, complementary to, or specifically hybridisable with, the corresponding part of a known, appropriate, mycolic acid bacterial sequence, such as the sequence shown in Fig. 4. Preferably parts of the sequence are used as probes, preferably of at least 100 nucleotides (but shorter sequences may be employed under high stringency conditions). The use of primers based on the sequence to screen and identify target sequences by PCR is also envisaged.

The identified putative inducible promoter can then be tested to see if it is operational as described in more detail below. Briefly, the putative promoter is provided in a vector upstream of a protein coding sequence (e.g. a reporter gene) at a position in which it is believed to be operatively linked to that coding sequence. A suitable host is transformed with the resulting vector. The presence or absence of the coding sequence expression product, in the presence of the inducing molecule, is determined. For putative transport proteins or catabolic enzymes identified by homology, function can be confirmed as described below.

As an alternative, or in addition to, homology screening, operon proteins which have catabolic enzymic activity can be screened for by their activity. For instance by contacting substrates for the enzymes (the analytes) with the host cells, or extracts therefrom, and observing for degradation products.

This approach can be used when the enzyme concerned may be successfully expressed in the recombinant host cell. For example, the R. corallina ohp operon was isolated by screening recombinant E. coli for expression of a catechol 2,3-dioxygenase activity induced in R. corallina when grown on monoaromatic compounds such as toluene. The substrate of this enzyme is catechol, a water soluble 2 hydroxyphenol which does not lyse E. coli.

In fact, R. corallina does express a mac catechol 2,3-dioxygenase activity in the presence of toluene. However that activity was not isolated in E. coli. Instead, the ohp-associated catechol 2,3-dioxygenase activity was isolated. This enzyme is induced by orthohydroxyphenylpropionic acid in the medium, although it does cleave catechol. A likely reason for the isolation of the ohp enzyme (rather than the mac one) is that functional screening in E. coli, even in those cases where it is possible, will depend not only on the requisite activity being expressed by the host, but also on the relative efficiency with which it is expressed. Thus using E. coli as the host, and using a broadly specific enzyme screen, those genes from nocardioform actinomycetes which are most efficiently expressed will be preferentially isolated.

Additionally, other potential substrates/analytes e.g. toluene are highly toxic to E. coli and may cause its membrane to destabilise leading to cell lysis. Further, gene isolation by function is limited to those genes that are expressed in the test bacterium. Because of their evolutionary distance from the mycolic acid

bacteria, established cloning hosts such as E. coli or Gram-positive bacteria such as Bacillus subtilis and Staphylococcus aureus may not effectively recognise mycolic acid bacterial gene regulatory signals and/or may not transport or survive in the presence or xenobiotics per se. Therefore, isolation by acquisition of novel-phenotype cannot easily be accomplished in these hosts.

In addition, when screening for proteins involved in binding or transporting the analyte, or transducing this binding event to the inducible promoter (e.g. transcription factors), it may be necessary to use a host in which other elements of the entire system (i.e. promoter and/or signal gene or catabolic enzymes) are present in order to demonstrate activity.

In order to circumvent these problems, in a most preferred embodiment of the second aspect, vectors comprising the inducible promoter and/or operon proteins are identified by means of a functional screen in a second host. This can avoid the difficulties described above. Preferably this second host is a suitable mycolic acid bacterium.

In order that the vectors can be maintained in the mycolic acid bacteria, they must encode replicons which can function in mycolic acid bacteria. These replicons can be those known in the art (e.g. based on characterised mycolic acid bacterial plasmids pSR1 (Batt et al., 1985). Alternatively the present inventors have provided a novel method of generating supercoiled or circular plasmid DNA from mycolic bacteria, and this method forms one part of the present invention. The diversity of the mycolic acid bacteria means that it is unlikely that a single replicon will be sufficient to construct biosensors in all strains encountered. Novel replicons which can be used either alone or in conjunction (two or more per vector) with other replicons to expand host range therefore provide a useful contribution to the art.

Thus, using the supercoiled/plasmid method of DNA isolation detailed in Example 4, two previously uncharacterised plasmids pRC100 and pRC158 have been discovered in soil mycolic acid bacteria Rhodococcus corallina and mycolic acid bacterium strain RC158 respectively.

Strain RC158 contains a supercoiled plasmid of approximately 14.57 kb. The plasmid, designated pRC158, contains at least five EcoRI restriction enzyme sites which can be used to digest the plasmid into a specific restriction pattern of five major restriction fragments of 4.3, 3.3, 2.9, 2.2 and 1.6 Kb DNA respectively. An approximately 100 kb plasmid, pRC100, was isolated from R. corallina

Replicons may be identified from novel plasmids by screening fragments obtained therefrom in disabled vectors containing marker proteins (for instance based on pJP7 described below) to see if they can replicate in mycolic acid bacteria.

Novel plasmids isolated using the method, and novel replicon elements isolated from them, form a fourth aspect of the present invention. These, and existing replicons, may be used to construct cloning vectors which replicate in several mycolic acid bacterial strains. Thus it is possible to clone, isolate by function and express specific genes from not only a single "type strain" as is the common practice in molecular biology but also in a variety of mycolic acid bacteria.

It is preferable that the transfer of the vectors comprising the putative inducible promoters and/or operon proteins to the second host (preferably mycolic acid bacteria) from the first host (preferably an established cloning systems such as E. coli) be achieved using bacterial conjugation. Experiments have shown that restriction enzyme activity in newly isolated mycolic acid bacteria effectively limits the efficiency of electroporation of incorrectly methylated plasmid DNA to

very low, or undetectable levels. It is known that most restriction enzymes preferentially act on double stranded DNA substrates. It is known that conjugative DNA transfer, however, involves a single-stranded DNA intermediate and is thus relatively immune to restriction. It is known that the IncPa conjugative plasmid RP4 can transfer its DNA into a wide range of bacteria by conjugation. Accordingly, a series of conjugatively mobilizable mycolic acid bacteria / E. coli shuttle vectors have been constructed by incorporation of a 440 bp region of the RP4 plasmid encoding the origin of transfer (pJP8 figure 1). Experiments have shown that RP4 oriT vectors can be complemented in trans for tra functions allowing conjugative mobilization into a variety of mycolic acid bacteria at high efficiency.

The vectors for use in the most preferred embodiment of second aspect of the invention (i.e. functional screening in a second host), themselves form a fifth aspect of the present invention, such vectors typically comprising:

(a) a replicon for mycolic acid bacteria
(b) a replicon for E. coli
(c) a conjugative origin of transfer
(d) a lambda cos site

An example of such a vector is that termed pJP8 (Figure 5). This comprises (a) pCY104oriV, (b) pBR322 oriV (c) RP4 oriT , and (d)a cos site; however it will be apparent to those skilled in the art that any of these could be substituted for a sequence having similar function, for instance substituting pRC100 or pRC158 minimal replicon sequences for the novel pCY104 replicon.

Further plasmids are pRV1 and pJH6 which comprise oriV (for replication in E. coli); oriT (for transfer); Kan (antibiotic marker); pSR1 (for replication); a cos site.

In use such vectors will further comprise a fragment containing the putative inducible promoter and/or operon

proteins and optionally a signal protein, such as have been described above.

Thus a gene library can be constructed in a mobilizable cosmid shuttle vector such as pJP8. After in vitro packaging, cosmids can be recovered by adsorption to E. coli carrying mcrABC mrr hsdSRM recA recO. Given the size of the mycolic acid genome (approximately 4 Mb) a 99% confidence gene library requires approximately 2500 colonies.

To screen for specific functions (either a complete reaction pathway or specific reactions) the packaged cosmids may be adsorbed to E. coli mcrABC mrr hsdSRM recA recO containing an IncP plasmid such as RK2. Since the RK2 plasmid encodes several antibiotic resistance genes, it is modified by random mutagenesis to disable antibiotic resistance genes which are also used as markers in the cosmid vector. From this transformed strain, the mobilizable cosmid shuttle vector may be conjugated into a wide variety of mycolic acid bacteria for functional screening. In any such screen, the choice of mycolic acid bacterial strain will be governed by the known catabolic functions of the strain. Thus entire pathways may be isolated by screening for gain of function. Alternatively, if a particular strain is known to require only one or a few catabolic activities these may be screened for by complementation.

Another novel shuttle vector, pRV1, can be recovered with high efficiency in a suitable E.coli host, and then transfer to a mycolic acid bacterial strain via conjugation (which minimises host restriction difficulties) for screening. Thus, in this embodiment, the E coli strain is just an interim host. Optionally conjugative systems can be put into place in this interim host to directly allow mating to follow phage adsorption, thus minimising the period in E.coli.

By incorporation of a signal gene adjacent to the cloning site in pJP8 or pRV1 used to construct the gene

library, transconjugant mycolic acid bacteria can be screened for inducible expression of a signal protein such as luciferase in the presence of specific molecules. This will rapidly isolate environmentally responsive promoter/operator/regulator elements.

Once identified, by any of the methods of the second aspect of the invention above, the putative inducible promoter and/or operon proteins may be modified by subcloning mutagenesis (typically within E. coli) and screened for enhanced function in mycolic acid bacteria.

The term 'modified' is used to mean a sequence obtainable by introducing changes into the full-length or part-length sequence, for example substitutions, insertions, and/or deletions. This may be achieved by any appropriate technique, including restriction of the sequence with an endonuclease followed by the insertion of a selected base sequence (using linkers if required) and ligation. Also possible is PCR-mediated mutagenesis using mutant primers.

It may, for instance, be preferable to add in or remove restriction sites in order to facilitate further cloning.

Alternatively, it may be particularly desirable to modify the binding protein/agent in order to modify its specificity and/or affinity for analyte.

Modified sequences according to the present invention may have a sequence at least 70% identical to the sequence of the full or part-length inducible promoter or operon protein as appropriate. Typically there is 80% or more, 90% or more 95% or more or 98% or more identity between the modified sequence and the authentic sequence. There may be up to five, for example up to ten or up to twenty or more nucleotide deletions, insertions and/or substitutions made to the full-length or part length sequence provided functionality is not totally lost.

Modified promoters and/or operon proteins can be

screened for functionality as described above in relation to isolating novel elements.

Nucleic acid encoding the authentic or modified promoter and/or genes encoding the operon proteins (plus such modified proteins themselves) identified or obtained by the method of the second aspect of the invention form a sixth aspect of the invention.

Thus one embodiment of the sixth aspect is the R. corallina ohp locus described in Figures 3 and 4 including the promoter and individual operon proteins encoding therein, and modifications thereof.

The authentic or modified promoter identified or obtained by the method of the second aspect of the invention may be used to inducibly express a heterologous signal protein in a transformed host; this use forms a seventh aspect of the present invention.

In one embodiment of the seventh aspect, there is disclosed a method of transforming a host with a vector encoding the inducible promoter as described above, operably linked to the signal gene (e.g. encoding luciferase).

The vector used in the seventh aspect may remain discrete in the host. Alternatively it may integrate into the genome of the host.

For a potential host (e.g. Corynebacterium) which does not express or generate the other components of the system which may be required to give biosensor function (for instance the operon proteins such as the transport protein to transport analyte into the cell; binding protein to bind analyte thereby inducing the promoter activity; cofactors required for signal protein activity etc.) these components can be added exogenously in order to perform the methods of the first aspect, or can be encoded on the vector used to introduce the inducible promoter or supplied *in trans* on a separate nucleic acid. Indeed, as stated above, any transport and binding proteins required for the functionality of the inducible

promoter will frequently form part the operon containing the promoter, and may thus be identified and isolated alongside it using the methods disclosed above.

Preferably, however, the host (e.g. a mycolic acid bacterium, either the same or different to that which provided the source of the inducible promoter, but preferably the same) will itself naturally express the other components of the system required to give biosensor function. This ensures all the required gene products for biosensor function are present.

Indeed in this latter case, the signal protein gene may be introduced into the host such that it is operably linked to an existing inducible promoter. In this embodiment of the seventh aspect of the invention the identification and or isolation of the promoter or associated proteins as described above ultimately provides the information required to allow targeting of the gene into this region. Typically this will be achieved by initiating targeted integration using aspects of the sequence forming part of the promoter region or operon.

Direct integration of a signal gene system such as luciferase (e.g. luxAB operon) into an environmentally responsive regulon in a mycolic acid containing bacterium may be more efficient than approaches based on isolation of gene(s) and its/their characterisation followed by construction of the biosensor. This integration can be achieved by transposition or by illegitimate or legitimate recombination between a genetic construct introduced into the cell and the target operon or gene cluster located on either the chromosome or an episomal element. In situations where a gene cluster or operon has been identified as above, by either screening in E. coli or direct functional cloning in a mycolic acid bacterium, site-specific recombination may be used to direct integration of the signal gene(s) (such as luciferase) into the regulon.

Vectors for use in the seventh aspect of the invention, form an eighth aspect of the invention. Such vectors will typically include: (a) the signal gene, plus (b) the inducible promoter, operably linked to the signal gene, or a sequence capable of initiating recombination of the signal gene such that it becomes operably linked with the inducible promoter. Further operon proteins (optionally modified) may also be included in the vector.

Vectors of the eighth aspect of the invention can be readily constructed on the basis of the present disclosure, for instance based on pJP7 (Figure 6) which is described in more detail below.

Strain derivatives encoding different gene dosage levels of the promoter/signal gene can be created by integration of the construct into the chromosome (low copy number/low sensitivity) or by use of medium or high copy number plasmids (medium or high sensitivity).

A ninth aspect of the invention is a (biosensor) host transformed with the vectors of the eighth aspect.

In using the transformants of the ninth aspect in the methods of the first aspect, the signal (such as bacterial luciferase) may be detected extracellularly using a photomultiplier or photodiode or any other photosensitive device. This maintains the cell integrity and thus resistance to environmental shock.

Also embraced within the scope of the present invention are kits for performing the various aspects of the invention. For instance a kit suitable for use in the first aspect may comprise a preparation of the microorganism, plus further means for carrying out the contact or observation steps e.g. buffers, co-factors (e.g. luciferin for addition to luciferase). A kit for performing the second aspect may include any of the following: selective buffer, a non-ionic detergent, any means for carrying out the screening process (e.g. primers, probes, substrates for catabolic enzymes, vectors for transfer into a second host). Kits for

performing the seventh aspect may include vectors for generating biosensors plus other means for transforming hosts with them (e.g. buffers etc.).

The invention will now be further described with reference to the following non-limiting Figures and Examples. Other embodiments falling within the scope of the present invention will occur to those skilled in the art in the light of these.

Figures

Figure 1 - shows an agarose gel on which digestions of the novel plasmid pRC100 has been run, as described in Example 5.

Figure 2 - shows an agarose gel on which digestions of the novel plasmid pRC158 has been run, as described in Example 5.

Figure 3 - shows a schematic view of the R. corallina ohp operon obtained by functional screening in E. coli, as described in Example 7. The schematic shows location of predicted genes: Regulator, Transport, Monooxygenase, Hydroxymuconic semialdehye hydrolase, Alcohol dehydrogenase. Initiator and terminator codons are shown as half height and full height lines respectively. Base coordinates refer to the Figure 4 sequence. The location of predicted promoter regions and direction are indicated by arrows. The molecular weights and coordinates of ohp genes are tabulated.

Figure 4 - shows the complete listing of the R. corallina ohp operon as described in Example 7. It includes a portion of a putative nitropropane promoter (5' of the regulator).

Figure 5 - shows a schematic diagram of the pJP8 vector of the present invention, as described in Example 8. Plasmid size is about 8.51 kb. pJP8 is a mycolic acid bacterium - E. coli mobilizable cosmid vector. It carries pCY104 replicon; is Kanamycin resistant 15 μg/ml mycolic acid bacteria, 50 μg/ml E. coli. It also carries lambda cos site, RP4 oriT site and a multiple cloning site.

Figure 6 - shows a schematic diagram of the pJP7 vector of the present invention, as described in Example 9. Plasmid size is about 10.66 kb. pJP7 is a mobilizable E. coli/Rhodococcus/Nocardia suicide/luciferase integration vector encoding luxAB signal genes, sacB gene

and thiostreppton resistance in Rhodococcus/Nocardia only up to 75 μg/ml (typically 1-10 μg/ml used in selections). The vector is RP4/RK2 mobilizable. By cloning a region of homology into the region upstream of the luxAB cassette, insertion can be targeted.

Figure 7 - shows a schematic diagram of the pRV1 vector of the present invention, as described in the Examples below. Plasmid pRV1 comprises a minimal pSR1 replicon (Archer & Sinskey, 1993 J Gen Microbiol 139: 1753-1759) which allows replication in C glutamicum. The pUC replication origin (Yanish et al, 1985 Gene 33: 103-119) allows replication in E. coli. Also included are a kanamycin resistance marker and the RP4 origin of conjugative transfer oriT. Transcription counterclockwise in the insert is terminated by the E. coli trpA terminator. Transcription clockwise into the insert may be initiated by the E. coli lac UV5 promoter.

Figure 8 - shows a schematic diagram of the pJH6 vector of the present invention, as described in the Examples below. This encodes the pSR1 replicon (*supra*) and the pBR322 replicon for replication in E. coli. Antiobiotic resistance markers are ampicillin (E. coli) and kanamycin (E. coli and mycolic acid bacteria). Transcription across the insert can be provided by exogenous expression of the T7 RNA polymerase (in vitro or in vivo).

Examples

Example 1 - A novel medium for oligotrophic screening

"MMRN" is prepared as a multicomponent stock to avoid the production of uncharacterised compounds during autoclaving. A "basic salts" stock is prepared containing 6g/L Na_2HPO_4; 3g/L KH_2PO_4; 1g/L NaCl; 4g/L $(NH_4)_2SO_4$; adjusted to pH 7.4 and made up to 989 mls with distilled water and autoclaved. A "100x A salts"

solution is prepared consisting of 20g/L $MgSO_4$; 2000 mg/L $FeSO_4.7H_2O$; 200 mg/L $FeCl_3$; 200 mg/L $MnSO_4.H_2O$ is prepared in distilled water and autoclaved. A "1000x B salts" solution consisting of 500 mg/L $ZnSO_4.7H_2O$; 200 mg/L $CuCl_2.2H_2O$; 200 mg/L $Na_2B_4O_7.10H_2O$; 100 mg/L $(NH_4)_7Mo_6O_{24}.4H_2O$ is prepared in distilled water and autoclaved. To prepare 1 litre of MMRN, sterile solutions of 989 mls basic salts, 10 mls 100 x A salts, 1 ml 1000 x B salts are combined. For solid media, agar is added to 1.4% w/v. Carbon-energy sources are supplemented to 80 μM final concentrations for soluble molecules, or as vapour for insoluble molecules (where their concentration is decided by their individual partition coefficients generally ranging from 3 to 40 μM). Petri plates or liquid cultures are incubated at 28°C to 30°C for up to 72 hours to accumulate sufficient biomass for genetic and biochemical testing.

Example 2 - Isolation of novel strains of mycolic acid containing bacteria from environmental samples using an oligotrophic screen and MMRN

Novel strains are a source of genetic diversity from which biosensors specific for particular xenobiotic compounds can be constructed. To isolate mycolic acid bacteria, for example Rhodococcus / Nocardia, from an environmental matrix such as soil, a rapid isolation technique is required. Isolation of bacteria from soil using standard laboratory media containing eutrophic levels of carbon preselects for eutrophic bacteria which can grow rapidly under these conditions. Oligotrophic bacteria such as Rhodococcus / Nocardia are rarely successfully isolated on such rich media. This can be carried out using MMRN to specifically enrich for and subsequently purify strains of mycolic acid-containing

bacteria which encode catabolic pathways whose expression is induced by a given xenobiotic. This methodology identifies molecules which are not only substrates, but are necessary and sufficient to induce the appropriate catabolic pathway. Soil suspensions from a matrix likely to express a desired phenotype (for instance a site known or believed to have been contaminated with a particular xenobiotic) can be used to inoculate MMRN supplemented with an oligotrophic level of a easily utilised carbon source (50μM). This provides an initial oligotrophic screen. Oligotrophic mycolic acid-containing bacteria are slow growing and may be expected to have formed colonies after 72 hours incubation at 28°C on MMRN paraffin. The incubation temperature appears to be highly selective of soil Nocardioform bacteria; Petri plates incubated at temperatures above 30°C fail to show detectable colonies. Colonies growing on alkanes can be initially screened for Nocardioform phenotype, selecting for crumbling, crenellated colonies, (possibly mucoid on rich media). Gram- and Ziehl-Neelsen-staining tests rapidly identify Gram-positive, mycolic acid-containing bacteria (Place a slide carrying a heat fixed film on a slide carrier over a sink. Flood with carbol fuchsin solution (basic fuchsin 5g; phenol, crystalline, 25g; 95% or absolute ethanol 50 ml; distilled water 500 ml) and heat until steam rises. Leave for 5 minutes, heating occasionally to keep the stain steaming. Wash with distilled water. Flood slide with 20% v/v sulphuric acid; wash off with distilled water, and repeat several times until the film is a faint pink. Finally wash with water. Treat with 95% v/v ethanol for 2 minutes. Wash with distilled water. Counterstain with 0.2% w/v malachite green. Wash and blot dry. Acid and alcohol fast organisms are red, other organisms are green).

Mycolic acid-containing bacteria may then be screened for specific hydrocarbon-inducible catabolic pathways using MMRN supplemented with the target

xenobiotic pollutant. Strains for which the target molecule is growth inducing may then be isolated and used to as a source of genetic regulatory elements for biosensors or as specific biocatalytic functions. Using this protocol mycolic acid containing bacteria have been and may be rapidly identified with novel and useful catabolic properties. This approach is also useful for identification and isolation of mycolic acid containing bacteria with biocatalytic properties.

Example 3 - Method for isolation of total DNA from mycolic acid bacteria

Bacterial strains were inoculated into 10 mls of MMRN supplemented with 500μM glucose 2% w/v L-glycine and incubated at 28°C for 30 to 40 hours. This medium supports relatively rapid growth of mycolic acid bacteria cells. The L-glycine present is misincorporated into peptidoglycan cell wall substantially weakening its resistance to osmotic shock (Katsumata, et al., 1984). Growth on MMRN appears to enhance the uptake of L-glycine and its apparent misincorporation into the cell arabinogalactan. During this growth phase, mycolic acid bacteria produce extensive surfactants which cause the accumulated biomass to clump into pellicles and exhibit a strong surface tension effect. These pellicles, which are highly resistant to lysozyme, may be broken up and the concentration of biosurfactants substantially reduced by washing the cell pellet in several culture volumes of 10 mM Tris pH8.0; 0.1% Tween 80 and finally resuspended in 1ml of 10 mM Tris HCl pH8.0, containing 10 mg/ml lysozyme. The lysozyme reaction is incubated 60 to 100 minutes at 37°C depending on the strain involved. Lysis is achieved by addition of 2% final (w/v) sodium dodecyl sulphate at 60°C 40 minutes. The nucleic acids are selectively purified from the cellular debris by sequential phenol, phenol: chloroform :isoamyl alcohol

(50:48:2 v/v) extractions. Nucleic acids are concentrated by ethanol precipitation in 2 M ammonium acetate. The nucleic acid pellet recovered is washed with 70% ethanol and resuspended in 100µl 10 mM Tris.HCl pH8.0, 1mM EDTA. 2 µl of this sample may be digested using restriction enzymes.

Example 4 - Method to isolate supercoiled/circular plasmid DNA from mycolic acid bacteria

50 mls Rhodococcus was cultured to mid-logarithmic phase in MMRN supplemented with 2% w/v L-glycine, 2% w/v D-glucose.

The cell pellet was washed in 10 mM Tris pH8.0 and 0.1% Tween 80. Resuspend cell pellet in 7.6 ml 6.7% sucrose; 50 mM Tris.HCl; 1 mM EDTA. Add 2 ml 40 mg/ml lysozyme in 10 mM Tris.HCl 1 mM EDTA. Incubate 37°C 15 minutes. Add 970 µl 250 mM EDTA, 50 mM Tris.HCl pH 8.0. Continue incubation for a further 105 minutes 37 °C. Lyse cells by addition of 600 µl 20% SDS 50 mM Tris.HCl, 20 mM EDTA pH 8.0. Incubate 55°C 30 minutes. Shear lysate by vigorous vortexing 30 seconds. Denature DNA by addition of 560 µl freshly prepared 3 M NaOH followed by gently mixing 10 minutes room temperature. Neutralise by addition of 1 ml 2.0 M Tris.HCl pH 7.0 with gentle mixing 10 minutes. Add 2.1 ml 20% SDS 50 mM Tris.HCl, 1 mM EDTA. Mix gently. Add 4.2 ml ice cold 5 M NaCl. Incubate on ice overnight or for several hours at least. Clear the cellular debris by centrifugation at 48000 g 4°C 90 minutes. The supernatant contains the DNA. Decant the supernatant by addition of an equal volume of ice cold isopropanol. Incubate -20°C 30 minutes. Pellet nucleic acids 4°C, 10000g 20 minutes.

Example 5: Novel plasmids and replicons obtained by the method of Example 4

Two multicopy plasmid replicons were isolated using the method of Example 4; pRC158 from strain RC158 and pRC100 from R. corallina.

Both plasmids have been digested with restriction enzymes to produce characteristic restriction patterns (Figures 1 and 2).

Plasmid pRC100, an approximately 100kb supercoiled circular plasmid present in R. corallina was prepared as described in the text. The agarose gel was loaded in lane 1 with Lambda DNA HindIII size markers (23,130 bp; 9,416 bp, 6,557 bp, 4,361 bp, 2,322 bp, 2,027 bp, 564 bp); lanes 2 to 9 inclusive were loaded with pRC100 digested with BamHI (5'GGATCC3'), Bc1I (5'TGATCA3'), Bg1II (5'AGATCT3'), EcoRI (5'GAATTC3'), HindIII (5'AAGCTT3'), KpnI 5'(GGTACC3'), SacI (5'GAGCTC3'), SalI (5'GTCGAC3') restriction endonuclease reactions which were carried out under standard conditions; lane 10 contains undigested (presumable supercoiled) pRC100 DNA; lane 11 pWW110/40121, lane 12 pWW110/4011; lane 13 pWW15/3202; lane 14 pUC18 lane 15 blank. The DNA fragments have been resolved on a 0.8% Agarose Tris-Acetate-EDTA gel. Southern blotting analysis using Gram-negative mono and polyaromatic catechol 2,3-dioxygenases failed to detect significant sequence conservation.

Plasmid pRC158 is a supercoiled plasmid of approximately 14.57 kb. The plasmid was digested with the EcoRI (5'GAATTC3') restriction endonuclease under standard conditions. The DNA fragments have been resolved on a 0.8% Agarose Tris-Acetate-EDTA gel. This pattern is unique and characteristic to pRC158. The plasmid contains at least five EcoRI restriction enzyme sites which can be used to digest the plasmid into a specific restriction pattern of five major restriction fragments of 4.3, 3.3, 2.9, 2.2 and 1.6 Kb DNA respectively.

These plasmids are relatively small, exhibit a high plasmid copy number and are easily isolated from

Rhodococcus / Nocardia. Therefore, they possess several characteristics which are suitable for the construction of Rhodococcus / Nocardia cloning vectors.

The DNA sequence of the minimal replicon regions of these plasmids may be determined by screening fragments obtained therefrom in disabled vectors containing marker proteins (for instance based on pJP7 described below) to see if they can replicate in mycolic acid bacteria.

Further plasmids e.g. pCY101 have also been isolated and sequenced using the methods of the present invention. The replicon from this plasmid was used in pJP8.

Example 6: Hybridisation screening for novel promoters and/or operon proteins

The test sample (host cells) are contacted with a nucleic acid molecule probe (preferably around 100 nucleotides or more) based on Figure 4 under suitable hybridisation conditions, and any test DNA which hybridises thereto is identified. Such screening is initially carried out under low-stringency conditions, which comprise a temperature of about 37°C or less, a formamide concentration of less than about 50%, and a moderate to low salt (e.g. Standard Saline Citrate ('SSC') = 0.15 M sodium chloride; 0.15 M sodium citrate; pH 7) concentration. Alternatively, a temperature of about 50°C or less and a high salt (e.g. 'SSPE'= 0.180 mM sodium chloride; 9 mM disodium hydrogen phosphate; 9 mM sodium dihydrogen phosphate; 1 mM sodium EDTA; pH 7.4). Preferably the screening is carried out at about 37°C, a form amide concentration of about 20%, and a salt concentration of about 5 X SSC, or a temperature of about 50°C and a salt concentration of about 2 X SSPE. These conditions will allow the identification of sequences which have a substantial degree of similarity with the probe sequence, without requiring the perfect homology for the identification of a stable hybrid. The phrase

'substantial similarity' refers to sequences which share at least 50% overall sequence identity. Preferably, hybridisation conditions will be selected which allow the identification of sequences having at least 70% sequence identity with the probe, while discriminating against sequences which have a lower level of sequence identity with respect to the probe. After low stringency hybridisation has been used to identify several clones having a substantial degree of similarity with the probe sequence, this subset of clones is then subjected to high stringency hybridisation, so as to identify those clones having a particularly high level of homology with respect to the probe sequences. High stringency conditions comprise a temperature of about 42°C or less, a form amide concentration of less than about 20%, and a low salt (SSC) concentration. Alternatively they may comprise a temperature of about 65°C or less, and a low salt (SSPE) concentration. Preferred conditions for such screening comprise a temperature of about 42°C, a form amide concentration of about 20%, and a salt concentration of about 2 X SSC, or a temperature of about 65°C, and a salt concentration of about 0.2 SSPE.

SUBE2 cont

Example 7 - Cloning aromatic degradative operon from Rhodococcus corallina by functional screening in E coli

To demonstrate the potential mycolic acid bacteria (e.g. Rhodococcus / Nocardia) have as biosensors and biocatalysts as well as to validate the novel genetic tools and approach to cloning of the present invention, a gene cluster or operon associated with aromatic degradation was cloned and isolated from Rhodococcus corallina. This gene cluster / operon appears to be a broad substrate range monoaromatic degradative pathway and has been designated monoaromatic catabolic (mac) gene cluster or operon. R. corallina was isolated from pristine soil in Canada and is an acknowledged

Rhodococcus type strain. This strain encodes a broad range of catabolic activities which include toluene, benzoate, phenol, cumine, cyamine. Genetic induction of the toluene degradative pathway in R. corallina occurs when toluene is supplied as vapour. This is a level of less than 200 ppm in water. Therefore, the sensitivity inherent in the biology of Rhodococcus is very close to those levels expected for biosensors in industrial use. Similar experiments using a naphthalene utilising Rhodococcus which is also supplied as a vapour

Biochemical assays of ring cleavage dioxygenase activities in crude enzyme extracts of R. corallina cells grown on MMRN supplemented with different growth-inducing xenobiotics indicated that the molecular specificity of ring cleavage dioxygenase induction is good. Toluene induced the meta pathway (although some ortho activity was observed) whereas benzoate and phenol exclusively induces the ortho pathway. Xylene, which is very closely related to toluene does not act as a growth inducing substrate. The closely related compounds toluene and benzoate but not xylene induce different ring-cleavage enzymes despite their relatively similar molecular shape. This behaviour and absence of induction with xylene suggests that the receptor for these or metabolites derived from these molecules is sensitive to minor electrostatic changes in their ligand. This strongly asserts that genetically constructed biosensors derived from these receptor molecules and their regulated promoter(s) will exhibit a level of specificity which exceeds that currently available as field test systems.

Since a clear catechol 2,3-dioxygenase activity was induced by toluene, but not by benzoate (indicating that the meta pathway in this strain is specifically induced by toluene), the catechol 2,3-dioxygenase activity can be used as a marker for gene(s), gene cluster(s) or operon(s) involved in its degradation.

The R. corallina catechol 2,3-dioxygenase structural

gene was isolated by functional screening of a partial Sau3A restriction enzyme digest-generated gene library in E. coli hsdRMmcrAB for using the commercially available cosmid cloning vector pWE15 (Wahl et al., 1987).

Because only a single enzyme activity has been used as a functional marker rather than complete acquisition of a phenotype and given the diversity of Rhodococcus / Nocardia metabolism and the genetic incompatibility between mycolic acid bacteria and E. coli it is possible that numerous catechol dioxygenases may exist but only some will be expressed successfully in E. coli. To facilitate expression of cloned DNA irrespective of the presence of an indigenous promoter element, a phage T7 promoter is located adjacent to the pWE15 unique BamHI restriction site into which the rhodococcal DNA was inserted. Phage T7 RNA polymerase (a single polypeptide) is supplied in trans from pGP1-2Sm. As a functional screen for 2,3-dioxygenase activity, catechol was sprayed onto nutrient agar plates supplemented with 15 μg/ml kanamycin, 50 μg/ml streptomycin, 0.1 mM isopropyl thiogalactoside (IPTG) incubated at 30°C to accumulate biomass. The expression of T7 polymerase is repressed by temperature sensitive phage lambda repressor which is itself expressed from an IPTG induced lacUV5 promoter. Thus incubation at 42°C leads to induction of T7 polymerase expression and so transcription of the pWE15 insert region from the T7 promoter (i.e. one direction of transcript alone).

Using the pGP1-2Sm T7 expression system, two colonies were isolated which encoded the characteristics catechol 2,3-dioxygenase activity from R. corallina. From approximately 3000 colonies of individual primary clones of R. corallina gene library in an E. coli hsdRMmcrAB strain, two colonies were observed to produce a deep yellow colour indicative of catechol 2,3-dioxygenase activity (2-hydroxymuconic semialdehyde) when exogenous catechol was supplied in phosphate buffer

(0.1M pH7.4). These clones were designated clone #1 and clone #2. Restriction enzyme mapping of both clone #1 and clone #2 DNA showed that both encode overlapping regions of DNA but were otherwise nonsibling clones; this is compatible with a primary screening of a cosmid library.

Southern blot analysis of R. corallina total cellular and plasmid DNA confirmed that the isolated catechol 2,3-dioxygenase locus in clones #1 and #2 are contiguous with an approximately 35 kb region R. corallina genomic DNA. The common region to both clones is comprised of seven major EcoRI restriction fragments (8.3, 7.2, 5.2, 4.9, 4.3, 2.4, 2.3 Kb respectively 34.6 kb in total). To confirm the continuity and source of the clone #1 and clone#2 inserts, an aliquot of clone #2 DNA, which contained a slightly longer R. corallina DNA insert than clone #1, was used as a source of DNA to synthesise a radioactive probe to identify homologous DNA restriction fragments present in an EcoRI restriction digest of total cellular R. corallina DNA as well as other bacterial DNA samples. An randomly picked pWE15 clone which did not express catechol 2,3-dioxygenase was chosen as one control (cosmid clone "clone # 4") and E. coli genomic DNA were selected as control DNAs. At the level of accuracy of the gel, the coincidence of the catechol 2,3-dioxygenase clones #1 and clone #2 DNA inserts relative to the genomic R. corallina EcoRI and SmaI restriction maps indicated that no gross deletions or rearrangements had occurred during the cloning. Significantly, there was no evidence for a supercoiled plasmid location for the catechol 2,3-dioxygenase gene indicating that the locus is chromosomally encoded (although pRC100 has been isolated from R. corallina (see Figure 1) this strain does not encode large linear plasmids). To investigate the potential for gene homologs to be identified a Rhodococcus strain RC161 which was isolated from North East England and so is

distinct from R. corallina (which also degrades toluene via meta cleavage but was isolated form soil in Canada) was included in the Southern Blot. There were three RC161 EcoRI restriction fragments which exhibited significant DNA sequence conservation with R. corallina sequences in clone #2. The nature of these sequences is under investigation.

Colony hybridisation to the R. corallina gene library secondary screen using the 2.4 Kb EcoRI restriction fragment of clone #2 as a source of radioactive probe identified four cosmid clones, pWE15#C, pWE15#D, pWE15#B and pWE15#G encoding overlapping regions of the R. corallina chromosome. Thus a region of the R.corallina genome with a contiguous length of approximately 70 kb has been cloned and isolated. These cosmids will provide a source of R. corallina DNA for future experiments.

The 35 Kb region encoded by clones #1 and #2 was mapped using four six base recognition restriction enzymes. An analysis of the map does not indicate inverted DNA map elements which could be consistent with a transposable element. This does not, however, preclude this possibility existing.

The sequence of the operon is described in Example 9 below.

Further plasmids which may be used for screening in accordance with the methods of the present invention are as follows:

pRV1

This is shown in Figure 7. It encodes the pSR1 replicon for Corynebacterium, the pUC replicon for E.coli, the RP4 oriT and a minimal cos PCR product. The multiple cloning site is under the control of the lac operon promoter allowing expression in E. coli.

The cos sequence in currently available in cosmids

such as pWE15 (Stratagene) and is encoded within an approximately 1 Kb region. However experiments showed that cos induced structural instability in several different plasmids. Analysis of the cos region in lambda suggested that the instability may be due to high levels of transcription entering the plasmid cos site and or transcription through adjacent lambda coding sequences which flank cos in the standard cosmid cloning vectors. To avoid problems with these extraneous elements, using computer-aided sequence analysis, the present inventors designed oligonucleotide primers to amplify the minimal cos element, free from flanking genes which may induce instability and occupy valuable cloning space. Additionally, experiments indicated that the cos PCR product induced structural instability in vectors carrying it. Therefore the cos PCR product was cloned into pRV1 (a preferred shuttle vector of the present invention) into a transcriptional quiet region of the plasmid. Transcription was blocked using a transcriptional terminator (trpA terminator from E. coli). This construct combines cosmid function with a mycolic acid replicon, an E.coli replicon, a selectable marker, a conjugative oriT, and a unique BamHI cloning site.

Briefly, the plasmid was prepared by cleaving plasmid pWSTIB (Peoples et al, 1988 Mol Microbiol 2(1): 63-72) with NheI and SalI to clone the *C glutamicum* replicon into the mobilisable plasmid pK19mob (Shäfer et al, 1994 Gene 145: 69-73) to form a shuttle vector designated pJH4. The minimal Cos site from wild-type phage (Promega) was amplified by PCR using primers which introduced two XbaI sites (5′ TCTAGA 3′) into the fragment.

The primers were:

F: 127 5′ CGCTGATTTGTATTGTCTG 3′ 145

R: 502 5' GACTTCCATTGTTCATTCC 3' 484

The fragment was cloned into pJH4 to give pRV1.

pJH6

This is shown in Figure 8. It also encodes the pSR1 replicon for Corynebacterium, the pUC replicon for E.coli, the RP4 oriT and a minimal cos PCR product. Inserted genes are expressed under the T3 and T7 promoters which are controlled by temperature shift, allowing the controlled production of genes which may impose a lethal phenotype.

Briefly, the plasmid was prepared by cleaving plasmid pWE15 (stratagene) with Agl III enzyme to remove unwanted SV40 ori and Neo sties. The NheI/BstBI fragment of pK18mob (Shäfer et al, 1994 Gene 145: 69-73) was cloned into pWE15-small to add a kanamycin resistance marker known to work in *C glutamicum* and *E coli*. The plasmid pWSTI B (above) was cleaved with BgIII and BamH1 enzymes to clone the pSR1 origin of replication of *C glutamicum* into pWE15-small. Finally RP4(OriT) was amplified by PCR using the following primers, which incorporate AatII restriction site:

F: 51171 5' AAAAGACGTCGGTGCGAATAAGGGACAGTG 3' 51190
R: 51395 5' AAAAGACGTCACAAAACAGCAGGGAAGCAG 3' 51376

The amplified fragment was cloned into the AatII site of the pWE15-small-Km-pSR1 construct to form the shuttled vector designated pJH6.

<u>Example 8 - A method for gene isolation from mycolic acid-containing bacteria by functional screening in Corynebacterium glutamicum</u>

A key aspect of this invention is the ability to genetically manipulate a variety of strains or species of

mycolic acid-containing bacteria such as Rhodococcus / Nocardia in a simple, effective way so as to clone and isolate gene(s), gene cluster(s) or operon(s) with applications as biosensors or biocatalysis.

The closely related mycolic acid-containing bacterium Corynebacterium glutamicum may be used as a host to express Rhodococcus / Nocardia genetic material. C. glutamicum shares a common cell wall type and probably similar genetic regulation to Rhodococcus / Nocardia but since it has been used extensively for the industrial production of amino acids and nucleotides it has lost or may never had encoded significant xenobiotic catabolic activity. It therefore represents a good "naïve" host to express Rhodococcus / Nocardia genes.

Restriction enzyme activity in natural isolates of Rhodococcus / Nocardia effectively limits the efficiency of electroporation to very low, or undetectable levels. Most restriction enzymes recognise double stranded DNA exclusively. Because single-stranded DNA is a necessary product of a replication fork, normal restriction enzyme activity in bacterial cells is limited to double stranded DNA substrates. Conjugative DNA transfer in Gram-negative, and most probably between Gram-positive bacteria as well, involves a single-stranded DNA intermediate. Conjugative DNA transfer should thus, generally, be relatively immune to restriction.

pJP8

The pJP8 plasmid may be used to introduce the library in the first host into a suitable mycolic acid bacterium such as corynebacterium or any mycolic acid bacterium which does not encode the desired phenotype.

The pJP8 plasmid is shown in Figure 5. The shuttle vector carries a approximately 400 bp region of the IncP RK2 conjugative plasmid which encodes the origin of transfer. This may be complemented in trans by IncP tra functions maintained on a suitable compatible recombinant

plasmid, or as an integrated construct in the host chromosome or by RK2 itself (modified to disrupt its kanamycin resistance gene - a marker used for pJP8).

Conjugation involves "effective contact" between the donor and recipient cells, which in this case are E coli encoding complementing tra functions and bearing the mobilizable cosmid vector and a suitable mycolic acid bacterium respectively. Effective contact is the formation of a cytoplasmic bridge between the two cells through which conjugative DNA transfer occurs. Thus donor and recipient cells are grown to mid to late logarithmic phase of growth in Lauria Bertini broth and MMRN supplemented with suitable carbon source at 37°C and 30°C respectively. Donor and recipient cells are washed in prewarmed media and mixed on a solid support matrix such as Lauria Bertini Agar plate and incubated at 37°C for up to 16 hours. The mating mixture is scraped from the plate and resuspended in 30°C Lauria Bertini broth, from which serial dilutions are prepared and plated on MMRN agar supplemented with drugs to counter select against the donor and recipient and select for the transconjugant mycolic acid bacterium. Commonly, naladixic acid selects against the donor and kanamycin resistance selects against the recipient. Thus, on a plates supplemented with both only the transconjugant may grow. The plates are incubated at 30°C for 40 hours.

Example 9 - DNA sequence of the proximal region of R. corallina ohp locus

The DNA sequence of approximately 7 Kb of R. corallina chromosomal DNA surrounding a catechol 2,3-dioxygenase has been determined using automated dye terminator sequencing reactions. A schematic of the current state of the data is presented in Figure 3 which shows at least seven genes which have been identified by protein sequence conservation with known protein motif

data (nitropropane dioxygenase, a putative regulatory protein orfR, monoaromatic monooxygenase, hydroxymuconic semialdehyde hydrolase, catechol 2,3-dioxygenase, alcohol dehydrogenase).

The sequence of this region in shown in Figure 4.

The predicated gene organisation of the ohp associated region is indicative of the presence of possibly two different catabolic gene clusters or operons; one involving the nitropropane dioxygenase the other the ohp gene cluster or operon. Such a genetic organisation suggests that a set of divergent promoter elements are located between the predicted regulatory gene orfR and the ohp monooxygenase structural gene. Similarly, another promoter could map immediately upstream of the divergent open reading frame which has conservation to nitropropane dioxygenase.

Example 10 - use of the promoter obtained in Example 9

The R. corallina genes identified by sequence conservation or by function are listed in Figure 3. These are potentially useful as catalytic functions in various chemical transformations. The regulatory protein associated with the putative ohp operon (possibly encoded by orfR) is involved in the control of transcriptional initiation at its target promoter. This regulatory protein encodes the specificity of the operon and as such is likely to be central to the biosensor function. Subcloning of the regulatory protein and its target promoter could permit novel biosensor activities to be introduced into other Rhodococcus /Nocardia strains. In addition, if this regulatory protein is subjected to mutagenesis, mutants with altered function could be identified (using a luciferase promoter probe driven by the regulated promoter). The regulatory protein has a specific capability to bind its ligand from the environment. It is therefore potentially useful as a

protein adsorbent for specific molecules. This could have application in analytical chemistry sample preparation.

An analysis of the 5' region of the predicted genes and the catechol 2,3-dioxygenase reading frame has allowed us to predict the sequence involved in translational initiation. These "ribosome binding sites" can be used as sequence guides or templates for the creation of synthetic oligonucleotides encoding functional Rhodococcus / Nocardia translational initiation sites. Mutagenesis of this region can identify potentially up and down regulating base sequences changes.

The ohp promoter region which controls expression of the cloned operon lies between two putative genes (orfR regulatory gene and orfT transport gene). In addition to forming the basis of a biosensor, the promoter and its cognate regulatory system also could be used as an inducible expression system for Rhodococcus / Nocardia and other mycolic acid-containing bacteria. The sequence of this region encodes the binding sites and regulatory elements or operators involved in control of the ohp and possibly other closely linked genes or operons. This region constitutes the first defined sequence for a Rhodococcus / Nocardia promoter region. It can be used as a probe to identify similar sequences within other mycolic acid containing bacteria such as Rhodococcus / Nocardia. This promoter sequence could be used as a region of homology to drive targeted recombination / insertion of signal gene(s) such as Vibrio luciferase.

A vector such as pJP7 (Figure 6) may be used as follows:

The vector is a 'suicide vector' which can be used to drive expression of bacterial luciferase genes in R. corallina. A portion of the ohp promoter region (Figure 4) is ligated into the unique pJP7 XbaI restriction site downstream of an E. coli trpA transcriptional terminator.

The sacB gene allows counter selection for the integrated plasmid thus selecting for a second cross-over within the plasmid sequences to produce a gene replacement of the wild type gene with an interrupted gene including luciferase. An aspect to this technique is the ability to introduce DNA constructs into the target cell in a hyperrecombinogenic, non-replicating form. Conjugatively mobilised plasmids may represent just such a form in that they may be single-stranded form. Thus the conjugatively mobilised plasmid pJP7 which cannot replicate in mycolic acid bacteria could be used directly to integrate DNA constructs into a wide range of mycolic acid bacterial strains.

Example 11 - Biosensor

The biosensor of the present invention is typically a recombinant mycolic acid containing bacteria which may be Rhodococcus / Nocardia cell. The natural gene-regulatory system which activates expression of catabolic gene(s), gene cluster(s) or operon(s) in response to the presence of specific class or type of inducing naturally-occurring or xenobiotic carbon substrate(s) has been genetically manipulated to induce the expression of some signal gene(s), such as (but not limited to) the Vibrio or Photobacterium bacterial luciferase in the presence of the inducer. This manipulation may have involved either incorporation of the signal gene(s) into a chromosomally- or episomally-encoded regulon under the control of a suitable environmentally-regulated promoter, or by direct sub-cloning of the regulated promoter to a rhodococcal / nocardial plasmid or other replicon or episomal element encoding a promoter-less signal gene(s). The genetic manipulation effecting the substitution or supplementation of the natural genes with the signal gene(s)may involve integration of the signal gene(s) gene

cluster(s) or operon into the host chromosome, plasmid or other episomal element so as to place it under inducible regulatory control or subcloning of the analyte (particularly hydrocarbon)-responsive promoter to a multicopy plasmid. The integration may involve site-specific recombination, transposition or illegitimate or homology-driven DNA recombination which is another aspect of this invention; however other methods of DNA integration such as the use of polymerase chain reaction (PCR) are not ruled out.

Signal to noise ratio can be readily improved in the recombinant system by enhancing or optimising expression or function of the signal gene, which may be luciferase, by means of improved gene translational signals and/or increasing levels of transcription by either raising transcriptional rates, mRNA stability or gene dosage of the construct (by subcloning to a plasmid or iterative gene integrations into a chromosome, plasmid or other episomal element). Thus, for instance, transcriptional efficiency of the luciferase genes luxAB can be increased by substitution of the Vibrio translational initiation signals with those from the ohp operon.

References

von der Osten et al (1989) Biotechnol Letts 11: 11-16.
Wahl et al (1987) Proc Natl Acad Sci 84: 2160-2164.
Katsumata et al (1989) J Bacteriol 159: 306-311.